

Annotated scientific text visualizer: design, development, and deployment

John Blake¹

Abstract. Prototypes of an annotated scientific text visualizer were designed, developed, and deployed. This pedagogic tool is designed to help undergraduates draft short research articles that conform to the generic expectations of their discourse community. This online tool enables users to discover and explore the language features present in short research articles. Users can select to visualize research articles in the field of computer science. The articles are categorized into four types. Users select to hide or reveal particular language features and their associated explanations in text, audio, or video formats. This enables them to create their own learning paths with this interactive tool. Students can use the visualizer to individualize their own learning interactively at their own pace on materials that are relevant to them.

Keywords: scientific writing, visualization, language features, individualized learning, discovery learning.

1. Introduction

This paper details the design, development, and deployment of a rapid prototype and an alpha release prototype of an annotated scientific text visualizer. This interactive pedagogic tool aims to help novice writers with English as an additional language by visualizing prototypical language features and providing multimedia explanations on demand. The inspiration for this tool stems from the noticing hypothesis (Schmidt, 2010), which claims that learners must first notice language features before learning them. Noticing is achieved through a discovery learning approach (Huang, 2008), enhanced using intelligent computer

1. University of Aizu, Aizu-wakamatsu, Japan; jblake@u-aizu.ac.jp; <https://orcid.org/0000-0002-3150-4995>

How to cite this article: Blake, J. (2019). Annotated scientific text visualizer: design, development, and deployment. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouéšny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 45-50). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.984>

assisted language learning (Amaral & Meurers, 2011). This is the first interactive visualization tool for novice writers of computer science research articles.

The purpose of this phase in the project was twofold. The first aim was to create a simple prototype to act as a visual aid that can be used in conjunction with the required specifications to show how the fully-fledged prototype should work. The second aim was to identify stakeholder expectations and improve our understanding of user needs to ensure that the scientific text visualizer not only meets but exceeds user expectations.

The following section provides the background details to the development of the scientific text visualizer. Section 3 describes the design specifications. Section 4 details the development of the rapid prototype, annotated articles, multimodal materials, and the alpha release. Section 5 concludes with the alpha release and lists future work.

2. Background

Both undergraduate and postgraduate students in the school of computer science at the University of Aizu in Northern Japan are required to submit short research articles in order to fulfill graduation requirements. This is a particularly onerous challenge for Japanese students who may have had little exposure to formal written English and less exposure to scientific writing. Ideally, students can dedicate a significant amount of time to read research articles in their field of research, and acquire the tacit knowledge required to write their own research paper. However, given the severe time constraints that many students face, this is not a viable option. A key problem for teachers of the associated technical writing courses is providing suitable examples and advice for all students. For example, within the field of computer science, some students may write more theoretical papers that rely on mathematical proofs while other students may develop and evaluate software, making it difficult for teachers to use examples that are relevant to all students.

Japanese students with little proficiency in English could make extensive use of Google Translate which since its switch to Google Neural Machine Translation now produces text that is more comprehensible than texts that those students could produce. This combined with the use of Grammarly or similar generic error detectors can produce somewhat comprehensible texts. This harnessing of technology however has no pedagogic purpose, and so the focus of this online tool

is to help writers learn more about the target genre by providing explanations on demand in the mode and medium that users prefer. This individualized automated support is both technically feasible and eminently scalable.

Individualized learning can solve the problem of differing needs and differing wants of students sharing the same class. Students can select example research articles which are most relevant to the type of research they are engaged in. They can then select the language features that they want to better understand. As this tool is online, writers can access it at will on any web-enabled device. This is particularly pertinent as many writers draft the final version of their graduation thesis over the New Year holiday period.

3. Design

A software requirements specification was created detailing use cases and requirements from the perspectives of students, teachers, and researchers. Research in computer science may be classified into four categories, namely empirical, experimental, practical, and theoretical. Once users select the type of research article, and the specific article itself, they can individualize their learning by showing or hiding various language features on demand. The features incorporated are listed in [Table 1](#). For each feature, explanations are provided in different modes (text, audio, and video) and mediums (Japanese and English).

Table 1. Language features to be visualized on demand using toggle buttons

#	Category	Details
1	organization	sections and moves
2	functions	description, explanation, exemplification, justification, reference to visuals
3	connections	coherence, cohesion (e.g. anaphoric and cataphoric pronouns)
4	linking	conjunctions, adverbs (transitions), and prepositions
5	tense	tense and aspect (e.g. present perfect progressive)
6	voice	passive, active (and ergative)
7	modality	hedging, boosting, and approximation
8	abstraction	packaging processes as nouns
9	information structure	end weight, focus, and flow (see Blake, 2015 for more details)
10	word type	first 1,000 words, second 2,000 words, academic word list

4. Development

Development can be divided into prototype creation (Prototype I and Prototype II) and materials creation (annotated articles and multimodal explanations). These are discussed in turn below.

4.1. Prototype I: Axure RP

A simple working prototype was made using Axure RP. A dropdown menu enables users to select research articles that are displayed in the center of the viewport. A row of ten toggle function buttons at the top allow users to hide/reveal language features. An exploratory panel appears above the research article when a function is selected. The exploratory panel contains an embedded video, a textual description, and a dropdown menu of other explanatory modes and mediums for the first function.

4.2. Materials creation

The initial dataset of 12 texts comprises abridged research articles written by undergraduates that were submitted as graduation theses, capstone projects, or final projects. Based on user feedback, texts over four pages were abridged. Where possible, raw text parsing is used, but when the state-of-the-art accuracy is insufficient, annotation tags are needed. For each function that requires annotation, html-like tags are used so that rule-based parsing can be used to visualize those particular language features. Users expect online learning resources to be interactive, highly visual and multimodal (Hafner, Chik, & Jones, 2015). Therefore, where possible, explanations are provided in text, image, audio, and file formats. Explanatory slideshows were created. Explanations were recorded in both English and Japanese to avoid the ‘L2 halting effect’ (Amaral & Meurers, 2011). The slideshows and audio files were merged to create videos.

4.3. Prototype II

The fully-fledged code version of the annotated text visualizer, Prototype II, allows users to select four types of computer science articles (practical, theoretical, empirical, and experimental) from a preloaded database of annotated articles. Users select the language features to be visualized on demand using toggle buttons to hide and reveal visualizations. When a toggle button is selected, the relevant features in the research article displayed in the viewport are highlighted and an explanatory panel appears. The explanatory panel is divided into two parts: embedded video

area and links to additional video, audio, or text explanations. Explanations are currently available in English or Japanese, but other languages may be added.

5. Discussion and conclusion

This pedagogic tool gives users the power to explore the form and function with visual, audio, and video explanations. Through exploring the visualizations and interacting with multimedia explanations, user awareness of generic expectations can be raised. This prototype tool is scalable and can be extended to deal with other scientific domains and different genres of writing.

The next phase of this three-year project is to extend the depth and breadth of the language features that can be visualized. The next version of the scientific text visualizer will be developed by a team of students using the Python Django web framework and Vue.js as the students have taken elective courses on these technologies. In contrast to the early prototypes, the next version will adopt a mobile-first approach.

6. Acknowledgments

This research was supported by a Japan Society for the Promotion of Science (JSPS) *Kakenhi* Grant-in-aid for Scientific Research entitled ‘Feature visualizer and detector for scientific texts’, Grant Number 19K00850. I would also like to thank Maxim Mozgovoy for his technical expertise.

References

- Amaral, L., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23(1), 4-24. <https://doi.org/10.1017/s0958344010000261>
- Blake, J. (2015). Incorporating information structure in the EAP curriculum. *Conference proceedings of 2nd International Symposium on Innovative Teaching and Research in ESP*. UEC, Tokyo. http://www.shilab.bunka.uec.ac.jp/esp2015/proceedings/Blake_ESP2015.pdf
- Hafner, C. A., Chik, A., & Jones, R. H. (2015). Digital literacies and language learning. *Language Learning & Technology*, 19(3), 1-7.
- Huang, L.-S. (2008). Using guided, corpus-aided discovery to generate active learning. *English Teaching Forum*, 46(4), 20-27.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan & I. Walker (Eds), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). National University of Singapore, Center for Language Studies.



Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2019 by Editors (collective work)
© 2019 by Authors (individual work)

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouéšny

Publication date: 2019/12/09

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2019.38.9782490057542>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover theme by © 2019 Frédéric Verolleman
Cover layout by © 2019 Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-54-2 (Ebook, PDF, colour)

ISBN13: 978-2-490057-55-9 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-53-5 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2019.